IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF MISSOURI
EASTERN DIVISION

A.O.A., *et al*.,	)	
	)	
Plaintiffs,	)	
	)	
vs.	)	Case No. 4:11-cv-00044-CDP
	)	(CONSOLIDATED)
THE DOE RUN RESOURCES CORPORATION, *et al*.,	)	
	)	
	)	
Defendants.	)	

PLAINTIFFS' REPLY IN SUPPORT OF THEIR MOTION TO EXCLUDE, IN PART, THE EXPERT OPINIONS OF DR. SHAHROKH ROUHANI

Table of Contents

Table of Authorities

Cases

Other Authorities

Rules

A discrete portion of Dr. Shahrokh Rouhani's March 19, 2021, supplemental report[1] contains an unreliable analysis of David Sullivan's air modeling. Rouhani – a statistician – analyzed Sullivan's three sets of modeling results by comparing hour-by-hour ***measured*** air concentrations of lead, arsenic, and sulfur dioxide recorded by Defendants at air monitoring sites in and around La Oroya at specific dates and times to ***modeled*** concentrations generated by Sullivan's CALPUFF air model ***at those same dates and times***.[2] This method of comparing modeled and measured air concentrations "paired in time and space" is not a reliable way to evaluate the validity of Sullivan's air model. Hour-by-hour changes in wind speed and direction, temperature, and industrial operations are impossible to predict accurately, and so Rouhani's methodology will almost always show poor model performance on an hourly basis.

Defendants point to academic literature that supposedly supports their position, but they misinterpret and misapply the literature they cite. They seek to bootstrap Rouhani's unreliable "concurrent comparisons" as a complement to his unchallenged nonconcurrent comparisons,[3] but "Daubert's standard of reliability 'extends to each step in an expert's analysis all the way through the step that connects the work of the expert to the particular case.'" *McLaughlin v. Bnsf Ry. Co.*, 439 F. Supp. 3d 1173, 1185 (D. Neb. 2020) (quoting *In re Paoli R.R. Yard Pcb Litig.*, 35 F.3d 717, 743 (3d Cir. 1994)). Defendants also misunderstand Sullivan's modeling. In sum, Defendants have failed to establish that Rouhani's concurrent comparisons are reliable and admissible under Rule 702 and *Daubert*.

[1] Doc. 1248-1.

[2] *See id*. at ¶¶ 22-25.

[3] Defendants devote several pages of their opposition to legally irrelevant discussions of aspects of Rouhani's testimony that Plaintiffs do not challenge. *See* Doc. 1281 at 4-5, 7-8, 12-13. While Plaintiffs disagree with claims made in those discussions, this reply does not address arguments that do not relate to the present motion.

Rouhani's "concurrent comparisons" should also be excluded under Rule 403. Sullivan has never claimed that his modeled concentrations are accurate on an hourly or even daily basis, and this case involves long-term exposures to lead, arsenic, and sulfur dioxide (among other toxic substances) over a period of many years. Whatever scant probative value Rouhani's concurrent comparisons might have is substantially outweighed by the danger of confusing and misleading the jury. The Court should enter an order precluding Rouhani from offering his concurrent comparison opinions at trial.

I. Argument

A. Rouhani's hourly and daily paired analysis is an unreliable and irrelevant method for evaluating Sullivan's long-term predictions.

Defendants have not met their burden of showing that Rouhani's concurrent comparison opinions are the product of reliable principles and methods. *See* Fed. R. Evid. 702(c). There are too many variables – such as wind speed and direction and daily changes in the Complex's operations – for any air model to predict concentrations accurately down to the hour. No experienced air modeler would evaluate a CALPUFF model on an hourly or even daily basis and expect it to show good model performance. Rouhani himself admits that, in his limited experience performing statistical analyses of air models, he has ***never*** found an air model that could accurately or reliably predict hourly or daily air quality at a particular location.[4] A methodology that ***always*** yields the same result has no predictive or analytical power and is unhelpful. *Cf. United States v. Watkins*, 10 F.4th 1179, 1184 (11th Cir. 2021) ("A standard or test that always produces the same result is not a standard or test. A box that will always be checked is not useful in an analysis.").

[4] *See* Doc. 1248-2, 4/14/2021 Deposition of Dr. Shahrokh Rouhani, at 82:17-23.

According to Defendants, Plaintiffs "can hardly complain" that Rouhani used daily and hourly results to perform his paired analyses because he "took the modeled and measured data directly from Mr. Sullivan's reliance materials provided to Defendants in December 2020."[5] But this argument misunderstands Sullivan's process. Doe Run Peru took readings of the ***air pollution concentrations*** in La Oroya every third day,[6] but Sullivan's air modeling is built from ***estimated monthly emissions rates***.[7]

To generate long-term air lead pollution predictions that could be compared to the measured air pollution data from every third day, Sullivan input estimated ***monthly*** lead emission rates to the CALPUFF modeling software. CALPUFF then incorporated measured meteorological data for the specific days that Defendants had air pollution measurements to generate predicted ***daily*** lead concentrations lining up with those same days.[8] Sullivan then averaged all his modeled daily concentrations for a given month to create a monthly average that forms the basis of his analysis (on annual, three-month, or distributional bases, for arsenic, lead, and sulfur dioxide, respectively).[9] These longer-term figures are the relevant and reliable output of Sullivan's modeling; creating daily modeled lead concentrations for days when Defendants had measured data was simply an intermediate step toward generating long-term concentration predictions.

---

[5] Doc. 1281 at 9-10.

[6] These intermittent measurements are shown in the middle column on Doc. 1281-5 ("Measured Daily Lead Concentrations"). The top of this column shows measured daily lead concentrations from the Hotel Inca, Huanchan, Casaracra, and Sindicato air monitoring sites for January 2, 5, 8, and 11, 2007, and so on. *Id*. at pdf p. 2.

[7] *See* Doc. 1248-6, Sullivan Rebuttal Rep. (May 28, 2021), at pdf p. 63.

[8] These modeled daily concentrations for every third day can be seen in the left column on Doc. 1281-5 ("Modeled Daily Lead Concentrations"). While this paragraph describes Sullivan's methodology with respect to lead, his methodologies for modeling arsenic and sulfur dioxide concentrations were functionally similar. *See* Docs. 1281-6 and 1281-7 (showing similar data for arsenic and sulfur dioxide, respectively).

[9] *See* Doc. 1248 at 5; *see also* Doc. 1248-3, Sullivan Rep. (Feb. 17, 2019), at pdf pp. 129-140; Doc. 1248-6 at pdf p. 54.

Sullivan's model was never intended to reliably predict daily or hourly concentrations; indeed, Sullivan never claimed that it would.[10] In other words, the question is not simply what sort of comparison is ***possible*** but rather what sort of verification is ***meaningful***. By way of analogy, no one can predict a specific sequence of outcomes from tossing a pair of dice, but one can test whether a pair of dice is fair over many rolls by comparing the characteristics of the distribution of outcomes (*e.g.*, mean, variance, frequency of occurrence of each outcome) with what we know eventually should occur. Here, Rouhani is criticizing Sullivan for failing to predict individual dice rolls. Sullivan's model assigns values to the individual dice rolls so that they can be aggregated to generate predictions. But its predictive value comes from its long-term averages. The accuracy of the modeled hourly concentrations is particularly irrelevant where, as here, the exposures at issue were continuous over many years.

B. The academic literature does not support Rouhani's unreliable methodology.

It is not generally accepted in the air modeling community to assess a model like Sullivan's by interpreting daily modeling predictions in isolation or expecting it to predict concentrations on a day-by-day (or hour-by-hour) basis. Indeed, Rouhani could not cite ***any*** scientific literature that supports his pairing in time and space methodology at his deposition.[11] Defendants now cite some academic literature in their opposition, but "[t]his approach smacks of post-hoc rationalization and is devoid of the intellectual rigor that *Daubert* demands. Put bluntly, this is not how good science is done." *Haller v. Astrazeneca Pharm. LP*, 598 F. Supp. 2d 1271, 1296-97 (M.D. Fla. 2009)); *see also* Fed. R. Civ. P. 26(a)(2)(B) (an expert's report "must contain

[10] *See* Doc. 1248 at 9; *see also* Doc. 1248-6 at pdf pp. 63-65.

[11] *See* Doc. 1248-2 at 88:21-89:24.

... a complete statement of all opinions the witness will express ***and the basis and reasons for them***") (emphasis added).

Contrary to Defendants' claims,[12] J.C. Chang & S.R. Hanna, *Air Quality Model Performance Evaluation*, 87 Meteorol. Atmos. Physics 167 (2004),[13] does not support Rouhani's concurrent comparison methodology.[14] That article concerns ***validation testing*** of models such as CALPUFF or AERMOD.[15] Validation testing involves collecting field data on uncomplicated terrain with research-grade meteorological equipment to confirm that a model is reliable where exact emissions sources, durations, and amounts are known.[16] In those circumstances, there can be an interest in comparisons paired in time and space.[17] Otherwise, "because of variations in wind direction," Chang and Hanna approvingly cite literature "stress[ing] the use of ... data unpaired in time or space" and state that "unpaired in space or time comparisons are usually sufficient."[18] Defendants also omit the authors' observation that air models "often ***completely fail***" when paired in time and space.[19]

---

[12] *See* Doc. 1281 at 1-2, 14-17.

[13] *See* Doc. 1248-7.

[14] Moreover, there is no indication that Rouhani relied on the paper in formulating his opinions. In fact, he testified at his deposition that he had never seen the article before it was handed to him. *See* Doc. 1248-2 at 137:13-138:19.

[15] *See, e.g.*, Doc. 1248-7 at pdf p. 2 (describing model verification and validation).

[16] *See id*. at pdf p. 19 ("Most of the reported model evaluation studies are associated with 'research grade' field experiments with good instruments and uncomplicated terrain and simple source scenarios."); *see also id*. at pdf p. 20 (describing modeling exercise in Salt Lake City involving "tracer studies," wherein "three SF_6 [sulfur hexafluoride] releases were made every two hours and the release duration was one hour from a [single point] source near street level[.]").

[17] The authors also suggest time and space pairing could be useful for modeling the dispersion of a chemical weapon released at a specific time and location. *See, e.g.*, *id*. at pdf p. 5 ("For military applications, the location of the dosage footprint of a chemical agent cloud is an important piece of information."); *see also id*. at pdf p. 19 (describing "applications such as emergency response [and] homeland security").

[18] *Id*. at pdf p. 19.

[19] Doc. 1281 at 16 (citing Doc. 1248-7 at pdf p. 7) (emphasis added).

Model validation based on precise tracer data emitted over flat terrain has little in common with Sullivan's task, which involved modeling concentrations of pollutants emitted from hundreds of release points, at varying rates and for changing durations, into a highly complex terrain using only estimated monthly emissions rates. Furthermore, the La Oroya Complex operated for many years, constantly exposing Plaintiffs to high concentrations or pollutants. There was no need to develop demonstrably accurate measurements at specific hours and locations. Sullivan did not endeavor to do so, and Chang and Hanna do not recommend it.

Defendants' other citations are similarly unhelpful to them on this issue. As the title of one exhibit[20] indicates, it is a validation study based on tracer data and thus, like the Salt Lake City study discussed by Chang and Hanna, has little in common with Sullivan's assignment.[21] Another study involved abundant measured pollutant data and simple terrain,[22] and even in that case, the paired analyses were remarkably inaccurate.[23] Defendants' final exhibit is a 44-year-old paper that does not mention pairing in time and space.[24] Defendants do not explain how the language they pull from that article offers any support for their arguments.[25]

---

[20] Doc. 1281-8, Arthur S. Rood, *Performance Evaluation of AERMOD, CALPUFF, and Legacy Air Dispersion Models Using the Winter Validation Tracer Study Dataset*, 89 Atmospheric Env't 707 (2014).

[21] *See id*. at pdf p. 2 ("The purpose of this paper is to examine the performance of [modeling software] using the Winter Validation Tracer Study (WVTS) data set …."); *see also supra* n.16.

[22] Doc. 1281-9, Kali D. Frost, *AERMOD Performance Evaluation for Three Coal-Fired Electrical Generating Units in Southwest Indiana*, 64 J. Air & Waste Mgmt. Ass'n (2014), at pdf p. 1. ("The sites are characterized by tall, buoyant stacks, flat terrain, multiple SO_2 monitors, and relatively isolated locations.").

[23] *Id*. ("Analysis … paired in time and space indicated poor model performance …. [A] scientific evaluation examining hourly paired monitor and model values at concentrations of interest indicates overprediction and underprediction bias that is outside of acceptable model performance measures.").

[24] Doc. 1281-10, U.S. Environmental Protection Agency, *Guideline on Air Quality Models,* EPA-450/2-78-027 (Apr. 1978). Meanwhile, Defendants chastise Chang and Hanna for referring to "the opinion of a different author in an article from thirty years ago[.]" Doc. 1281 at 16-17.

[25] *See* Doc. 1281 at 18.

Academic literature "is not intended to represent or replace the standard of care by which the adequacy of a given professional service must be judged, nor should it be applied without consideration of a project's many unique aspects."[26] Sullivan is the only air modeler in this case. As he has explained, "[a]ny experienced and objective air quality modeler could have told [Rouhani] that it would be highly unlikely that any correlation would be shown when constraining an air quality model in time and space[.]"[27] But Rouhani is not an air modeling expert, and as a result, he incorporates a methodology that is both meaningless and unhelpful.[28] *See Shipp v. Murphy*, 9 F.4th 694, 701 (8th Cir. 2021) (internal marks omitted) ("Rule 702 does require that the area of the witness's competence match the subject matter of the witness's testimony.").

C. Defendants' other justifications for Rouhani's unreliable paired analyses are unavailing.

None of Defendants' other arguments salvage Rouhani's unreliable methodology. Defendants attempt to justify Rouhani's concurrent comparisons by pointing to his ***nonconcurrent*** comparisons and claiming that this other analysis provides cover for the rest of his opinions.[29] Their argument is a version of "the more, the better."[30] But pointing to opinions that Plaintiffs have not moved to exclude does not fix what is wrong with the unreliable analysis. *See In re Johnson & Johnson Talcum Powder Prods. Mktg., Sales Practices & Prods. Litig.*, 509

[26] Exh. 8, ASTM Int'l, *Standard Guide for Statistical Performance of Air Quality Dispersion Models*, D6589-05 (Apr. 2015) at pdf p. 1.

[27] Doc. 1248-6 at pdf p. 54.

[28] *See* Doc. 1248 at 5-6.

[29] *See, e.g.*, Doc. 1281 at 5 ("To better understand Mr. Sullivan's model…, Dr. Rouhani also conducted a more comprehensive analysis that included concurrent comparisons …."); *id*. at 12 ("Plaintiffs have failed to recognize Dr. Rouhani's broader statistical analysis …. Rouhani actually began his analysis by conducting a nonconcurrent comparison ….").

[30] *See id*. at 14 (Rouhani "conducted a more comprehensive evaluation that included both concurrent and nonconcurrent comparisons") (quotations omitted).

F. Supp. 3d 116, 139-40 (D.N.J. 2020) (quoting *In re Pfizer Inc. Securities Litig.*, 819 F.3d 642, 665 (2d Cir. 2016)) (a district court may "exclude unreliable portions of an expert's testimony '[w]hen faced with ... testimony that contains both reliable and unreliable opinions"); *cf. Heller v. Shaw Indus., Inc.*, 167 F.3d 146, 159 n.8 (3d Cir. 1999) (recognizing that unreliable portion of expert's testimony must be excluded even where other portion of expert's testimony does not merit exclusion). Defendants suggest that it can be useful to present several different performance measures,[31] but "any step the expert has taken that renders the analysis unreliable renders the expert's testimony inadmissible[.]" *Weisen v. N. Tier Retail LLC*, No. 19-CV-2624, 2021 WL 2661507, at *4 (D. Minn. June 29, 2021).[32]

Defendants assert that Rouhani's concurrent comparisons are appropriate responses to Sullivan's claims about the reliability of his modeling results. For instance, they argue that Rouhani has used a variety of statistical analyses to examine whether Sullivan's claims about the general reliability of his model's results are verifiable under a wide variety of circumstances.[33] But Defendants cannot explain why any purported flaws in ***Sullivan's*** analysis justify admitting ***Rouhani's*** unreliable testimony. *See Redd v. Depuy Orthopaedics*, 700 F. App'x 551, 554 (8th Cir. 2017) ("The proponent of the expert testimony must prove its admissibility by a preponderance of the evidence.").

Defendants also misinterpret Sullivan's opinions. They focus on his December 2020 report (and particularly Table E-5 therein), which discusses the Complex's emissions and release

[31] *See, e.g.*, Doc. 1281 at 14 ("[T]he recommendation to use a variety of different performance measures … is a central and repeated theme of their article.").

[32] Doc. 1248-7 at pdf p. 7 ("The relative advantages of each performance measure are partly determined by the characteristics and distributions of the model predictions and observations[.]").

[33] *See* Doc. 1281 at 11-12 (initial caps removed) ("Dr. Rouhani's statistical comparisons respond directly to Mr. Sullivan's generic, unexplained, and unsupported statistical claims."); *see also id*. at 12 (("As compared to Mr. Sullivan's conclusory allegations, Dr. Rouhani's quantitative analyses constitute far more than sufficiently reliable responses for purposes of Rule 702."); *id*. at 3, 4, 6, 9, 11, 12, 19.

specifications, not air modeling.[34] The relevant analysis is in Sullivan's February 17, 2019, report,[35] which Defendants dismiss out of hand.[36] In that report, Sullivan explains that "average modeled concentrations ... were generally within a factor of +/- two of the average measured concentrations but with most of the modeled results ***lower than*** comparable measured concentrations (typically about half of the measured concentrations)."[37] In other words, the differences between Sullivan's modeling and Defendants' measured data are favorable to Defendants.[38] Sullivan supports his conclusion with a 12-page Appendix E to his February 2019 report, which contains comparisons of modeled and measured concentrations from 2001, 2006, 2007, and 2008,[39] and statistical analysis of his model's performance. In Tables E-13 through E-15, Sullivan shows regressions of his modeling with R-squared values of .915 (lead), .857 (arsenic), and .783 (sulfur dioxide), indicating strong predictive power.[40]

---

[34] *See id*. at 3, 4, 6, 9, 11, 12, 14, 15, 19; *see also generally* Doc. 1248-5 (titled "Review of Emissions and Release Specifications for the La Oroya Metallurgical Complex").

[35] *See* Doc. 1248-3 (titled "Air Quality & Meteorological Analysis of the Doe Run Peru Smelter and Refinery Operations in the Vicinity of La Oroya, Peru"), at pdf pp. 86, 129-140.

[36] *See* Doc. 1281 at 15. Defendants suggest that Sullivan only modeled his own reconstructed emissions inventory, *see id*. at 3, and claim that Sullivan's February 2019 modeling results are "now obsolete and based on old emission estimates." *Id.* at 15. But neither assertion is true. As explained in Plaintiffs' previous briefing, Sullivan ran his model three times to account for the parties' three different emissions inventories. *See* Doc. 1248 at 5, Doc. 1273 at 4. Indeed, Rouhani's Opinion 3 compares the daily and hourly modeled values of all three of Sullivan's model runs (using the three emissions inventories) against one another. *See* Doc. 1248-1 at ¶¶ 26-28 and Attach. C and D. No inventory is "obsolete"; the various model runs simply reflect different sets of assumptions, all of which show extreme environmental degradation. *See* Doc. 1248-3 at pdf pp. 40-88, 93-141, 160-174.

[37] Doc. 1248-3 at pdf p. 86 (emphasis added). Sullivan goes on to state on that same page of his 2019 report that "a standard rule of thumb for expected model performance is ~ +/- a factor of two," refuting Defendants' claim that "only later, in his 2021 Rebuttal Report, did Mr. Sullivan even reference his target accuracy goal of plus or minus a factor of two." Doc. 1281 at 3 (quotation omitted).

[38] *See* Doc. 1281 at 5 (Table 5 showing almost all differences as negative numbers). Defendants draw attention to Sullivan's modeling at the Sindicato Station. *See id*. at 4 n.1, 13. Plaintiffs explain in their opposition to Defendants' motion to exclude Sullivan that Sindicato's inappropriate siting led to measured concentrations that were too low and incorporate herein that discussion. *See* Doc. 1273 at 19-20.

[39] In their only mention of Sullivan's 2019 report, Defendants incorrectly claim that "the modeling predictions cited in [Sullivan's] 2019 Report were for the year 2006[.]" Doc. 1281 at 15.

[40] Doc. 1248-3 at pdf pp. 138-40. Defendants repeatedly claim Sullivan "fail[s] his own factor-of-two test," *see* Doc. 1281 at 3-6, 15-16, but Plaintiffs explain Defendants' mischaracterization of this rule of thumb as a hard-and-fast

II. Conclusion

Plaintiffs' targeted motion seeks exclusion of a discrete portion of one of Rouhani's reports because his methodology is unreliable and his opinions are likely to waste the jury's time and cause confusion by presenting scientifically meaningless testimony. "Analysts should understand the nuances in the data they are analyzing," but Rouhani "was simply processing numbers without understanding the context for his analyses."[41] Pursuant to Rules 702 and 403, the Court should exercise its gatekeeping role and exclude Rouhani's concurrent comparison opinions, as set forth in paragraphs 22-28 of his March 19, 2021, supplemental report.

Respectfully submitted,

SCHLICHTER BOGARD & DENTON LLP

By: /s/ *Nathan D. Stump*
Jerome J. Schlichter, #32225
Nelson G. Wolff, #40796
Kristine K. Kraft, #37971
Nathan D. Stump, #71641
100 South 4th Street, Suite 1200
St. Louis, Missouri 63102
Telephone: (314) 621-6115
jshlichter@uselaws.com
nwolff@uselaws.com
kkraft@uselaws.com
nstump@uselaws.com

Attorneys for Plaintiffs

rule in their opposition to Defendants' motion to exclude Sullivan and incorporate herein that discussion. *See* Doc. 1273 at 17-19. In any event, the accuracy of Sullivan's results goes to their weight, not admissibility.

[41] Doc. 1248-6 at pdf pp. 64-65.